

[SEQUENCE LISTING]

SEQUENCE ID No.: 1

LENGTH: 251

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: peptide

SEQUENCE:

Met Ser Met Pro Asn Ile Val Pro Pro Ala Glu Val Arg Thr Glu Gly  
5 10 15  
Leu Ser Leu Glu Glu Tyr Asp Glu Glu Gln Val Arg Leu Met Glu Glu  
20 25 30  
Arg Cys Ile Leu Val Asn Pro Asp Asp Val Ala Tyr Gly Glu Ala Ser  
35 40 45  
Lys Lys Thr Cys His Leu Met Ser Asp Ile Asn Ala Pro Lys Asp Leu  
50 55 60  
Leu His Arg Ala Phe Ser Val Phe Leu Phe Arg Pro Ser Asp Gly Ala  
65 70 75 80  
Leu Leu Leu Gln Arg Arg Ala Asp Glu Lys Ile Thr Phe Pro Gly Met  
85 90 95  
Trp Thr Asn Thr Cys Cys Ser His Pro Leu Ser Ile Lys Gly Glu Val  
100 105 110  
Glu Glu Glu Asn Gln Ile Gly Val Arg Arg Ala Ala Ser Arg Lys Leu  
115 120 125  
Glu His Glu Leu Gly Val Pro Thr Ser Ser Thr Pro Pro Asp Ser Phe  
130 135 140  
Thr Tyr Leu Thr Arg Ile His Tyr Leu Ala Pro Ser Asp Gly Leu Trp  
145 150 155 160  
Gly Glu His Glu Ile Asp Tyr Ile Leu Phe Ser Thr Thr Pro Thr Glu  
165 170 175  
His Thr Gly Asn Pro Asn Glu Val Ser Asp Thr Arg Tyr Val Thr Lys  
180 185 190  
Pro Glu Leu Gln Ala Met Phe Glu Asp Glu Ser Asn Ser Phe Thr Pro  
195 200 205  
Trp Phe Lys Leu Ile Ala Arg Asp Phe Leu Phe Gly Trp Trp Asp Gln  
210 215 220  
Leu Leu Ala Arg Arg Asn Glu Lys Gly Glu Val Asp Ala Lys Ser Leu  
225 230 235 240

Glu Asp Leu Ser Asp Asn Lys Val Trp Lys Met \*\*\*  
245 250

SEQUENCE ID No.: 2

SEQUENCE TYPE: amino acid

MOLECULAR TYPE: peptide

Met	Gln	Leu	Leu	Ala	Glu	Asp	Arg	Thr	Asp	His	Met	Arg	Gly	Ala	Ser
				5					10					15	
Thr	Trp	Ala	Gly	Gly	Gln	Ser	Gln	Asp	Glu	Leu	Met	Leu	Lys	Asp	Glu
			20					25					30		
Cys	Ile	Leu	Val	Asp	Ala	Asp	Asp	Asn	Ile	Thr	Gly	His	Val	Ser	Lys
		35				40						45			
Leu	Glu	Cys	His	Lys	Phe	Leu	Pro	His	Gln	Pro	Ala	Gly	Leu	Leu	His
	50					55					60				
Arg	Ala	Phe	Ser	Val	Phe	Leu	Phe	Asp	Asp	Gln	Gly	Arg	Leu	Leu	Leu
65					70					75					80
Gln	Gln	Arg	Ala	Arg	Ser	Lys	Ile	Thr	Phe	Pro	Ser	Val	Trp	Thr	Asn
				85					90					95	
Thr	Cys	Cys	Ser	His	Pro	Leu	His	Gly	Gln	Thr	Pro	Asp	Glu	Val	Asp
			100					105					110		
Gln	Leu	Ser	Gln	Val	Ala	Asp	Gly	Thr	Val	Pro	Gly	Ala	Lys	Ala	Ala
		115					120					125			
Ala	Ile	Arg	Lys	Leu	Glu	His	Glu	Leu	Gly	Ile	Pro	Ala	His	Gln	Leu
	130					135					140				
Pro	Ala	Ser	Ala	Phe	Arg	Phe	Leu	Thr	Arg	Leu	His	Tyr	Cys	Ala	Ala
145					150					155					160
Asp	Val	Gln	Pro	Ala	Ala	Thr	Gln	Ser	Ala	Leu	Trp	Gly	Glu	His	Glu
			165						170					175	
Met	Asp	Tyr	Ile	Leu	Phe	Ile	Arg	Ala	Asn	Val	Thr	Leu	Ala	Pro	Asn
		180						185					190		
Pro	Asp	Glu	Val	Asp	Glu	Val	Arg	Tyr	Val	Thr	Gln	Glu	Glu	Leu	Arg
		195					200					205			

Gln Met Met Gln Pro Asp Asn Gly Leu Gln Trp Ser Pro Trp Phe Arg  
 210 215 220  
 Ile Ile Ala Ala Arg Phe Leu Glu Arg Trp Trp Ala Asp Leu Asp Ala  
 225 230 235 240  
 Ala Leu Asn Thr Asp Lys His Glu Asp Trp Gly Thr Val His His Ile  
 245 250 255  
 Asn Glu Ala \*\*\*

SEQUENCE ID No.: 3

LENGTH: 288

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: peptide

SEQUENCE:

Met Thr Ala Asp Asn Asn Ser Met Pro His Gly Ala Val Ser Ser Tyr  
 5 10 15  
 Ala Lys Leu Val Gln Asn Gln Thr Pro Glu Asp Ile Leu Glu Glu Phe  
 20 25 30  
 Pro Glu Ile Ile Pro Leu Gln Gln Arg Pro Asn Thr Arg Ser Ser Glu  
 35 40 45  
 Thr Ser Asn Asp Glu Ser Gly Glu Thr Cys Phe Ser Gly His Asp Glu  
 50 55 60  
 Glu Gln Ile Lys Leu Met Asn Glu Asn Cys Ile Val Leu Asp Trp Asp  
 65 70 75 80  
 Asp Asn Ala Ile Gly Ala Gly Thr Lys Lys Val Cys His Leu Met Glu  
 85 90 95  
 Asn Ile Glu Lys Gly Leu Leu His Arg Ala Phe Ser Val Phe Ile Phe  
 100 105 110  
 Asn Glu Gln Gly Glu Leu Leu Leu Gln Gln Arg Ala Thr Glu Lys Ile  
 115 120 125  
 Thr Phe Pro Asp Leu Trp Thr Asn Thr Cys Cys Ser His Pro Leu Cys  
 130 135 140  
 Ile Asp Asp Glu Leu Gly Leu Lys Gly Lys Leu Asp Asp Lys Ile Lys  
 145 150 155 160  
 Gly Ala Ile Thr Ala Ala Val Arg Lys Leu Asp His Glu Leu Gly Ile

					165						170					175			
Pro	Glu	Asp	Glu	Thr	Lys	Thr	Arg	Gly	Lys	Phe	His	Phe	Leu	Asn	Arg				
			180						185				190						
Ile	His	Tyr	Met	Ala	Pro	Ser	Asn	Glu	Pro	Trp	Gly	Glu	His	Glu	Ile				
		195					200					205							
Asp	Tyr	Ile	Leu	Phe	Tyr	Lys	Ile	Asn	Ala	Lys	Glu	Asn	Leu	Thr	Val				
		210				215					220								
Asn	Pro	Asn	Val	Asn	Glu	Val	Arg	Asp	Phe	Lys	Trp	Val	Ser	Pro	Asn				
225					230					235					240				
Asp	Leu	Lys	Thr	Met	Phe	Ala	Asp	Pro	Ser	Tyr	Lys	Phe	Thr	Pro	Trp				
			245						250					255					
Phe	Lys	Ile	Ile	Cys	Glu	Asn	Tyr	Leu	Phe	Asn	Trp	Trp	Glu	Gln	Leu				
		260					265						270						
Asp	Asp	Leu	Ser	Glu	Val	Glu	Asn	Asp	Arg	Gln	Ile	His	Arg	Met	Leu				
		275					280					285							

\*\*\*

SEQUENCE ID No.: 4

LENGTH: 1099

SEQUENCE TYPE: nucleic acid

STRANDNESS: double

TOPOLOGY: linear

MOLECULAR TYPE: cDNA

ORIGIN

ORGANISM: Phaffia rhodozyma

STRAIN: ATCC 24230

SEQUENCE CHARACTERISTIC

CHARACTERISTIC CODE: CDS

LOCATIION: 99..851

DETERMINATION METHOD: E

SEQUENCE:

CCCACGCGTC CGCACATCTC GCATATATCA CTTTCCTCCT TCCAGAACAA GTTCTGAGTC 60

AACCGAAAAG AAAGAAGGCA GAGGAAAATA TATTCTAG ATG TCC ATG CCC AAC ATT 116

Met Ser Met Pro Asn Ile

5

GTT CCC CCC GCC GAG GTC CGA ACC GAA GGA CTC AGT TTA GAA GAG TAC 164

Val Pro Pro Ala Glu Val Arg Thr Glu Gly Leu Ser Leu Glu Glu Tyr

10

15

20

GAT GAG GAG CAG GTC AGG CTG ATG GAG GAG CGA TGT ATT CTT GTT AAC 212

Asp Glu Glu Gln Val Arg Leu Met Glu Glu Arg Cys Ile Leu Val Asn

25

30

35

CCG GAC GAT GTG GCC TAT GGA GAG GCT TCG AAA AAG ACC TGC CAC TTG 260

Pro Asp Asp Val Ala Tyr Gly Glu Ala Ser Lys Lys Thr Cys His Leu

40

45

50

ATG TCC AAC ATC AAC GCG CCC AAG GAC CTC CTC CAC CGA GCA TTC TCC 308

Met Ser Asn Ile Asn Ala Pro Lys Asp Leu Leu His Arg Ala Phe Ser

55

60

65

70

GTG TTT CTC TTC CGC CCA TCG GAC GGA GCA CTC CTG CTT CAG CGA AGA 356

Val Phe Leu Phe Arg Pro Ser Asp Gly Ala Leu Leu Leu Gln Arg Arg

75

80

85

GCG GAC GAG AAG ATT ACG TTC CCT GGA ATG TGG ACC AAC ACG TGT TGC 404

Ala Asp Glu Lys Ile Thr Phe Pro Gly Met Trp Thr Asn Thr Cys Cys

90

95

100

AGT CAT CCT TTG AGC ATC AAG GGC GAG GTT GAA GAG GAG AAC CAG ATC 452

Ser His Pro Leu Ser Ile Lys Gly Glu Val Glu Glu Glu Asn Gln Ile

105

110

115

GGT GTT CGA CGA GCT GCG TCC CGA AAG TTG GAG CAC GAG CTT GGC GTG 500

Gly Val Arg Arg Ala Ala Ser Arg Lys Leu Glu His Glu Leu Gly Val

120

125

130

CCT ACA TCG TCG ACT CCG CCC GAC TCG TTC ACC TAC CTC ACT AGG ATA 548

Pro Thr Ser Ser Thr Pro Pro Asp Ser Phe Thr Tyr Leu Thr Arg Ile

135

140

145

150

CAT TAC CTC GCT CCG AGT GAC GGA CTC TGG GGA GAA CAC GAG ATC GAC 596

His Tyr Leu Ala Pro Ser Asp Gly Leu Trp Gly Glu His Glu Ile Asp

155

160

165

TAC ATT CTC TTC TCA ACC ACA CCT ACA GAA CAC ACT GGA AAC CCT AAC 644

Tyr Ile Leu Phe Ser Thr Thr Pro Thr Glu His Thr Gly Asn Pro Asn

170

175

180

GAA GTC TCT GAC ACT CGA TAT GTC ACC AAG CCC GAG CTC CAG GCG ATG 692

Glu Val Ser Asp Thr Arg Tyr Val Thr Lys Pro Glu Leu Gln Ala Met

96211 "GTE/EBD"

185	190	195	
TTT GAG GAC GAG TCT AAC TCA	TTT ACC CCT TGG TTC AAG TTG ATT GCC	740	
Phe Glu Asp Glu Ser Asn Ser	Phe Thr Pro Trp Phe Lys Leu Ile Ala		
200	205	210	
CGA GAC TTC CTG TTT GGC TGG TGG GAT CAA CTT CTC GCC AGA CGA AAT	788		
Arg Asp Phe Leu Phe Gly Trp Trp Asp Gln Leu Leu Ala Arg Arg Asn			
215	220	225	230
GAA AAG GGT GAG GTC GAT GCC AAA TCG TTG GAG GAT CTC TCG GAC AAC	836		
Glu Lys Gly Glu Val Asp Ala Lys Ser Leu Glu Asp Leu Ser Asp Asn			
235	240	245	
AAA GTC TGG AAG ATG TAGTCGACC CTTCTTTCTG TACAGTCATC TCAGTTCGCC	890		
Lys Val Trp Lys Met ***			
250			
TGTTGGTTGC TTGCTTCTTG CTCTTCTTTC TATATATCTT TTTTCTTGCC TGGGTAGACT	950		
TGATCTTTCT ACATAGCATA CGCATAACATA CATAAACTCT ATTTCTTGTT CTTTATCTCT	1010		
CTTCTAAGGG AATCTTCAAG ATCAATTTCT TTTTGGGCTA CAACATTTC AATCAATGTT	1070		
GCTTTTCAGA CTACAAAAAA AAAAAAAAAA	1099		

SEQUENCE ID No.: 5

LENGTH: 1074

SEQUENCE TYPE: nucleic acid

STRANDNESS: double

TOPOLOGY: linear

MOLECULAR TYPE: cDNA

ORIGIN

ORGANISM: Haematococcus pluvialis

STRAIN: NIES-144

SEQUENCE CHARACTERISTIC



TTG CAC TAC TGC GCC GCG GAC GTG CAG CCG GCT GCG ACA CAA TCA GCA 654  
 Leu His Tyr Cys Ala Ala Asp Val Gln Pro Ala Ala Thr Gln Ser Ala  
 155 160 165 170  
 CTC TGG GGC GAG CAC GAA ATG GAC TAC ATC TTA TTC ATC CGG GCC AAC 702  
 Leu Trp Gly Glu His Glu Met Asp Tyr Ile Leu Phe Ile Arg Ala Asn  
 175 180 185  
 GTC ACC CTT GCG CCC AAC CCT GAC GAG GTG GAC GAA GTC AGG TAC GTG 750  
 Val Thr Leu Ala Pro Asn Pro Asp Glu Val Asp Glu Val Arg Tyr Val  
 190 195 200  
 ACG CAG GAG GAG CTG CGG CAG ATG ATG CAG CCG GAC AAT GGG TTG CAA 798  
 Thr Gln Glu Glu Leu Arg Gln Met Met Gln Pro Asp Asn Gly Leu Gln  
 205 210 215  
 TGG TCG CCG TGG TTT CGC ATC ATC GCC GCG CGC TTC CTT GAG CGC TGG 846  
 Trp Ser Pro Trp Phe Arg Ile Ile Ala Ala Arg Phe Leu Glu Arg Trp  
 220 225 230  
 TGG GCT GAC CTA GAC GCG GCC CTG AAC ACT GAC AAA CAC GAG GAT TGG 894  
 Trp Ala Asp Leu Asp Ala Ala Leu Asn Thr Asp Lys His Glu Asp Trp  
 235 240 245 250  
 GGA ACG GTG CAT CAC ATC AAC GAA GCG TGA AAACAG AAGCTGTAGG 940  
 Gly Thr Val His His Ile Asn Glu Ala \*\*\*  
 255  
 ATGTCAAGAC ACGTCATGAG GGGGCTTGGC ATCTTGCGCG CTTCGTATCT CTTTTTACTG 1000  
 AGACTGAACC TGCAGCTGGA GACAATGGTG AGCCCAATTC AACTTTCCGC TGCACTGGAA 1060  
 AAAAAAAAAA AAAA 1074

SEQUENCE ID No.: 6

LENGTH: 1058

SEQUENCE TYPE: nucleic acid

STRANDNESS: double

TOPOLOGY: linear

MOLECULAR TYPE: genomic DNA

ORIGIN



ORGANISM: *Saccharomyces cerevisiae*

STRAIN: S288C

SEQUENCE CHARACTERISTIC

CHARACTERISTIC CODE: CDS

LOCATION: 187..1050

DETERMINATION METHOD: S

SEQUENCE:

TCGATGGGGG TTGCCTTTCT TTTTCGGTCT TAACTCCATT TATATTTATT TATTCATTTT 60  
TATCTATTTA ACAGGAAACA GTTTTCTAGT GACAAGAAGG CGTATATCCC ACTTAATTCA 120  
ATATTAGAGT ATTCGTATTT GGAATACAGG AAGAGTAAAA ATAAGCCAAA AATTCATTAC 180  
ACCTCA ATG ACT GCC GAC AAC AAT AGT ATG CCC CAT GGT GCA GTA TCT AGT 231  
Met Thr Ala Asp Asn Asn Ser Met Pro His Gly Ala Val Ser Ser  
5 10 15  
TAC GCC AAA TTA GTG CAA AAC CAA ACA CCT GAA GAC ATT TTG GAA GAG 279  
Tyr Ala Lys Leu Val Gln Asn Gln Thr Pro Glu Asp Ile Leu Glu Glu  
20 25 30  
TTT CCT GAA ATT ATT CCA TTA CAA CAA AGA CCT AAT ACC CGA TCT AGT 327  
Phe Pro Glu Ile Ile Pro Leu Gln Gln Arg Pro Asn Thr Arg Ser Ser  
35 40 45  
GAG ACG TCA AAT GAC GAA AGC GGA GAA ACA TGT TTT TCT GGT CAT GAT 375  
Glu Thr Ser Asn Asp Glu Ser Gly Glu Thr Cys Phe Ser Gly His Asp  
50 55 60  
GAG GAG CAA ATT AAG TTA ATG AAT GAA AAT TGT ATT GTT TTG GAT TGG 423  
Glu Glu Gln Ile Lys Leu Met Asn Glu Asn Cys Ile Val Leu Asp Trp  
65 70 75  
GAC GAT AAT GCT ATT GGT GCC GGT ACC AAG AAA GTT TGT CAT TTA ATG 471  
Asp Asp Asn Ala Ile Gly Ala Gly Thr Lys Lys Val Cys His Leu Met  
80 85 90 95  
GAA AAT ATT GAA AAG GGT TTA CTA CAT CGT GCA TTC TCC GTC TTT ATT 519  
Glu Asn Ile Glu Lys Gly Leu Leu His Arg Ala Phe Ser Val Phe Ile  
100 105 110  
TTC AAT GAA CAA GGT GAA TTA CTT TTA CAA CAA AGA GCC ACT GAA AAA 567  
Phe Asn Glu Gln Gly Glu Leu Leu Leu Gln Gln Arg Ala Thr Glu Lys  
115 120 125

ATA	ACT	TTC	CCT	GAT	CTT	TGG	ACT	AAC	ACA	TGC	TGC	TCT	CAT	CCA	CTA	615
Ile	Thr	Phe	Pro	Asp	Leu	Trp	Thr	Asn	Thr	Cys	Cys	Ser	His	Pro	Leu	
		130					135					140				
TGT	ATT	GAT	GAC	GAA	TTA	GGT	TTG	AAG	GGT	AAG	CTA	GAC	GAT	AAG	ATT	663
Cys	Ile	Asp	Asp	Glu	Leu	Gly	Leu	Lys	Gly	Lys	Leu	Asp	Asp	Lys	Ile	
		145				150					155					
AAG	GGC	GCT	ATT	ACT	GCG	GCG	GTG	AGA	AAA	CTA	GAT	CAT	GAA	TTA	GGT	711
Lys	Gly	Ala	Ile	Thr	Ala	Ala	Val	Arg	Lys	Leu	Asp	His	Glu	Leu	Gly	
160					165					170					175	
ATT	CCA	GAA	GAT	GAA	ACT	AAG	ACA	AGG	GGT	AAG	TTT	CAC	TTT	TTA	AAC	759
Ile	Pro	Glu	Asp	Glu	Thr	Lys	Thr	Arg	Gly	Lys	Phe	His	Phe	Leu	Asn	
				180				185						190		
AGA	ATC	CAT	TAC	ATG	GCA	CCA	AGC	AAT	GAA	CCA	TGG	GGT	GAA	CAT	GAA	807
Arg	Ile	His	Tyr	Met	Ala	Pro	Ser	Asn	Glu	Pro	Trp	Gly	Glu	His	Glu	
			195					200					205			
ATT	GAT	TAC	ATC	CTA	TTT	TAT	AAG	ATC	AAC	GCT	AAA	GAA	AAC	TTG	ACT	855
Ile	Asp	Tyr	Ile	Leu	Phe	Tyr	Lys	Ile	Asn	Ala	Lys	Glu	Asn	Leu	Thr	
		210					215					220				
GTC	AAC	CCA	AAC	GTC	AAT	GAA	GTT	AGA	GAC	TTC	AAA	TGG	GTT	TCA	CCA	903
Val	Asn	Pro	Asn	Val	Asn	Glu	Val	Arg	Asp	Phe	Lys	Trp	Val	Ser	Pro	
		225				230					235					
AAT	GAT	TTG	AAA	ACT	ATG	TTT	GCT	GAC	CCA	AGT	TAC	AAG	TTT	ACG	CCT	951
Asn	Asp	Leu	Lys	Thr	Met	Phe	Ala	Asp	Pro	Ser	Tyr	Lys	Phe	Thr	Pro	
240					245					250					255	
TGG	TTT	AAG	ATT	ATT	TGC	GAG	AAT	TAC	TTA	TTC	AAC	TGG	TGG	GAG	CAA	999
Trp	Phe	Lys	Ile	Ile	Cys	Glu	Asn	Tyr	Leu	Phe	Asn	Trp	Trp	Glu	Gln	
				260				265						270		
TA	GAT	GAC	CTT	TCT	GAA	GTG	GAA	AAT	GAC	AGG	CAA	ATT	CAT	AGA	ATG	1047
Leu	Asp	Asp	Leu	Ser	Glu	Val	Glu	Asn	Asp	Arg	Gln	Ile	His	Arg	Met	
				275				280						285		
CTA	TAA	CAACG														
Leu	***															